

AMENDMENTS TO THE SPECIFICATION

Please amend the paragraph beginning on page 2, line 10, to read as follows:

On the basis of the fact that many organisms which properly grow at a temperature of 80° or above are located at the root of a phylogenetic tree by 16S rRNA (Fig. 1) shown by Woese et al., the inventors had an idea that the ancestors common to eubacteria, eukaryotes and archaebacteria might be ultra-thermophilic bacteria. On the basis of this supposition, the inventors have gotten an idea that although protein of many kinds of existing thermophilic bacteria are not always the protein of a true ancestral protein, the proteins having an amino acid sequence of the ancestral or an amino acid sequence close to the ancestral sequence might have a further improved thermostability. The inventors have completed the present invention on the basis of an idea that for designing and producing a thermostable protein, it is more important that the amino acid sequence of ancestral protein is estimated and mimicked than that only the sequence and the higher-order structure of protein of a thermophilic bacterium are analyzed and mimicked.

Please amend the paragraph beginning on page 2, line 23, to read as follows:

Namely, the present invention provides a method for improving thermostability of proteins, which comprises the steps of

- (i) comparing amino acid sequences of proteins derived belonging to the same family and deriving from two or more species which, wherein said proteins evolutionarily correspond to each other in a phylogenetic tree;
- (ii) estimating an amino acid sequence of an ancestral protein corresponding to the amino acid sequences compared in step (i); and,
- (iii) and comparing the amino acid residues in the amino acid sequence in one of the proteins compared in step (i) with amino acid residues at a corresponding position in the ancestral protein estimated in step (ii), and replacing one or more of the amino acid residues different

B2
from those of the ancestral protein with the same amino acid residues as those of the ancestral protein.

Please amend the paragraph beginning on page 3, line 4, to read as follows:

The present invention may further comprise the ~~setps~~ steps of

B3
(iv) testing the proteins obtained in step (iii) for thermostability; and
(v) selecting a protein having improved thermostability.

Please amend the paragraph beginning on page 6, line 16, to read as follows:

B4
Amino acid sequences of protein to be improved in thermostability can be also obtained from an already known ~~data-base~~ database. When an amino acid sequence is to be newly determined, any method for determining amino acid sequence known in the art can be employed. It is also possible to estimate the amino acid sequence by obtaining a nucleic acid encoding the protein according to the information of partial amino acid sequence, determining the nucleic acid sequence by a well-known sequencing techniques and estimating the amino acid sequence from the nucleic acid sequence.

Please amend the paragraph beginning on page 6, line 16, to read as follows:

B5
After the multiple alignment of the obtained amino acid sequences from the species, the amino acid sequences obtained from the respective species are compared with each other. Some methods for the multiple alignment are known. One of the methods is based on the ~~maximum~~ maximum parsimony principle for minimizing the change due to the insertion, deletion, replacement, etc. Computer programs for implementing this principle have been developed, which can be used or available. For example, TreeAlign is known among them. From DDBJ, "malign" which is the 1990 version of the program can be used. Because species ~~which~~ that are evolutionarily close to each other in the phylogenetic tree are selected in the present invention, phylogenetic information has already been utilized in the multiple alignment and, as a result, the alignment is more suitable than that in a case of no

phylogenetic information can be conducted. Information from at least three species is utilized for the multiple alignment. The larger the number of origin of the data to be used for the alignment, the more suitable the information. Furthermore, each of the species to be compared preferably contains one or more thermophilic bacteria or archaebacteria, based on the aforementioned reason. It is also preferred that it contains a family protein, namely another protein expected to be derived from the same ancestral protein.

Please amend the paragraph beginning on page 7, line 16, to read as follows:

After obtaining the results of the alignment, amino acid sequence of the ancestral protein can be estimated on the phylogenetic tree. For this purpose, the maximum parsimony method or maximal likelihood method is utilizable. The procedure of such a method is well known to those skilled in the art [see, for example, Young, Z., Kumar, S and Nei. M, Genetics 141, 1641-16510, 1995; Steward, C. -B. Active ancestral molecules, Nature 374, 12-13, 1995; and ~~Molecular~~ Evolutionary Molecular Evolutionary Genetics, Columbia University Press, New York, USA, 1987]. For example, the maximal parsimony method which can be employed in the present invention is, in short, a method wherein an ancestral type having the minimal number of the mutation expected to occur after the estimation of the ancestral type is likely estimated to be the true ancestral type. The maximal likelihood method can be employed instead of the maximum parsimony method. Also, a program PROTPARS (included in PHYLIP) for directly estimating the ancestral type from the amino acid sequence according to the maximum parsimony method can be also employed. Because the phylogenetic tree and ancestral amino acid are principally estimated at the same time in those methods, it is not always necessary to prepare the phylogenetic tree when such a method is employed. However, the preparation of the phylogenetic tree is preferred particularly when the ancestral amino acid is to be estimated by manual calculation. The ancestral amino acid sequence can be determined by the following maximum parsimony

method or maximal likelihood method according to a phylogenetic tree produced by the above-described method or another already known method, particularly based on an already published phylogenetic tree.

36
Cont
Please amend the paragraph beginning on page 8, line 16, to read as follows:

37
Ancestral amino acids in respective sites of the multiply aligned residues can be determined by means of a phylogenetic tree obtained by any method. For example, Fig. 4 shows amino acid residues from various organisms corresponding to residue 152 of *Sulfolobus* sp. strain 7 of IPMDH. Amino acids at this position in the organisms shown in Fig. 4 are R, S, K or E. When both residues in species close to each other in the phylogenetic tree are R, it can be estimated that in the ancestral species common to them (shown by the binding point connecting two species in the phylogenetic tree), the amino acid residue corresponding to residue 152 of *Sulfolobus* sp. strain 7 would be R for the following reasons:

When R is the ancestral type, only one variation can elucidate the mechanism of the realization occurrence of the amino acid residue corresponding to residue 152 of *Sulfolobus* sp. strain 7 in the present species, while when S is the ancestral type, two or more times of variation must be taken into consideration.

38
Please amend the paragraph beginning on page 11, line 15, to read as follows:

The molecular biological techniques such as introduction of a gene into a host, cloning of genes and site-specific mutagenesis including ung⁺ hosts, are well known by those skilled in the art. For these techniques, for example, Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, and F. M. Ausubel et al. (eds) al. (Eds), Current Protocols in Molecular Biology, John Wiley & Sons, Inc. (1994) can be referred to. Further, kits for carrying out these molecular biological techniques are commercially available. The mutation thus introduced can be confirmed by determining the nucleotide sequence. When a

38
restriction enzyme site has been introduced in the primer for the variation introduction, the introduction of the mutation can be more easily confirmed on the basis of the fact that it can be digested by a corresponding restriction enzyme.

Please amend the paragraph beginning on page 17, Table 1, to read as follows:

Table 1

Multiple alignment of amino acid sequences of IPMDH and ICDH

Enzyme and species	Partial amino acid sequence
IPMDH	89 97 150 158 256 263 280 285
<i>Sulfolobus</i> sp. strain □	YDMYANIRP---IAKVG-LNFA---VHGAAFDI---MMYERM
<i>Thermus thermophilus</i>	QDLFANLRP---VARVA-FEAA---VHGSAPDI---MMLEHA
<i>Bacillus subtilis</i>	LDLFANLRP---VIREG-FKMA---VHGSAPDI---MLLRTS
<i>Escherichia coli</i>	FKLFSNLRP---IARIA-FESA---AGGSAPDI---LLRYS
<i>Agrobacterium tumefaciens</i>	LELFANLRP---IASVA-FELA---VHGSAPDI---MCLRYS
<i>Saccharomyces cerevisiae</i>	LQLYANLRP---ITRMAAF-MA---CHGSAPDL---MMLKLS
<i>Neurospora crassa</i>	LGTYGNLRP---IARLAGF-LA---IHGSAPDI---MMLRYS
ICDH	89 97 150 158 256 263 280 285
<i>Saccharomyces cerevisiae</i>	FGLFANVRP---VIRYA-FEYA---VHGSAPDI---MMLNHM
<i>Bos Taurus</i> (3/4)	FDLYANVRP---IAEFA-FEYA---VHGTAPDI---MMLRHM
<i>Bacillus subtilis</i>	LDLFVCLRP---LVRAA-IDYA---THGTAPKY---LLLEHL
<i>Escherichia coli</i>	LDLYICLRP---LVRAA-IEYA---THGTAPKY---MMLRHM
<u>Ancestral species</u>	xDLxANLRP---IARxAxFExA---VHGSAPDI---MMLxxx
<u>Ancestral species</u> (predicted)	
modified amino acids and their positions	L L R A S P L <a region> <b region> <c region> <d region> b' b''

NE
See above
Please amend the paragraph beginning on page 11, line 15, to read as follows:

After annealing, 2 μ l of 10 x synthetic buffer (50 mM Tris-HCl, 20 mM MgCl₂, 5 mM dNTPs, 10 mM ATP, 20 mM DTT, pH 7.9), 1 μ l of T4 DNA ligase and 1 μ l of T4 DNA polymerase were added to the annealed solution. The obtained mixture was kept in ice for 5 minutes and then at room temperature for 5 minutes, and then incubated at 37°C for 90 minutes. 4 μ l of the reaction mixture was taken and mixed with 100 μ l of *Escherichia coli* MC 1061 competent cells. The obtained mixture was left to stand at 0°C for 20 minutes, at 42°C for 1 minute and 0°C for 2 minutes. 450 μ l of 2xYT medium was added thereto and they were left to stand at 37°C for 1 hour. 138.5 μ l of the culture liquid was poured into 5 ml of 2xYT liquid medium containing 100 μ g /ml of ampicillin. After overnight culture, the plasmid DNA was recovered from the cells by alkali-SDS method.

Please amend the paragraph beginning on page 20, line 12, to read as follows:

B10
Because ancestral variant in abcd region was not obtained, however, this variant was constructed from the ancestral ancestral a region variant and ancestral bcd region variant.

Please amend the paragraph beginning on page 20, line 15, to read as follows:

B11
Ancestral bcd region variant plasmid pE7-SB21bcd DNA obtained as described above was digested with *Sma* I. On the other hand, a variant plasmid pE7-SB21a DNA was digested with *Xba* I and *Eco* RI, and DNA segment encoding the intended enzyme was subcloned into *Xba* I – *Eco* RI multicloning site of pUC118 to obtain plasmid pUC118-SB21a. pUC118-SB21a was digested with *Sma* I and ligated with the above-described bcd region ancestral variant plasmid DNA digested with *Sma* I to obtain pUC118-SB21abcd. Then pUC118-SB21abcd and pE7-SB21 were digested with *Xba* I and *Eco* RI. They were mixed together to obtain expression plasmid pE7-SB21abcd for the ancestral variant in abcd region.

Please amend the paragraph beginning on page 23, line 15, to read as follows:

B12
The half-lives of natural and variant (ancestral) enzymes at 99°C were determined as follows: Enzyme solutions having a protein concentration of 0.25 mg/ml (for b', b'', b, c and d variants) or 1.0 mg/ml (for abcd variant) were prepared by using a potassium phosphate buffer (20mM KHPO₄, 0.5 mM EDTA, 1 mM DTT, pH 7.0). Also for natural IPMDH, enzyme solutions having protein concentrations of 0.25 mg/ml and 1.0 mg/ml were prepared. These enzyme solutions were heat-treated at 99°C for 10, 20, 30, 60 or 120 minutes. After the completion of the treatment, the enzyme solutions were left to stand in ice for 5 minutes and then centrifuged at 12,000 rpm at 4°C for 20 minutes. The supernatant was recovered from each product. 10 µl of each supernatant was used to determine the activity at 75°C. The determination was repeatedly conducted 3 times for each sample, and the average of results was taken as the residual activity. The residual activity was plotted in a graph wherein the horizontal axis ~~represent~~ represents the time, and the ordinates represent the relative activity (time 0 was represented as 100). The time at which the relative activity was 50 % was taken as the half-life T_{1/2}. At the same time, the specific activity was also determined. The results are shown in Tables 3 and 4.

Please amend the paragraph beginning on page 25, line 9, to read as follows:

B13
Wild type IPMDH from *Thermus thermophilus* and ancestral IPMDH were purified using the similar procedure as described in Example 2, making it a proviso that the third nucleotide of several codons of the gene were changed to A or T to ~~lager~~ enhance the production of the protein, because IPMDH gene from *Thermus thermophilus* is GC rich, which may decrease the expression of the gene. The final yields from 1L culture were 184mg/L for wild type, 11.3mg/L for ancestral variant F53L and 8.4mg/L for ancestral variant V181T.

Please amend the paragraph beginning on page 26, line 10, to read as follows:

The reason why the thermostability of F53L variant was reduced to less than the thermostability of wild type may reside in the following factors: Investigation of the amino acid sequence around residue 53 revealed that the residue 58 in *Thermus thermophilus* ^{B14} *thermophilus* is Arg, while it is Leu or Val in many other species. From the fact, it is believed that the structure became unstable by changing the amino acid residue at position 53 to Leu which cannot fill the space between the residue 53 and Arg at position 58, unlike Phe, and the thermostability was reduced as a result.

Please amend the paragraph beginning on page 27, line 2, to read as follows:

Amino acid sequences of IPMDH from representative species and ICDH from various species were obtained from NCBI database and they were subjected to the multiple alignment using Clustal X, ~~an~~ a software for alignment (Fig. 14). Also the composite phylogenetic tree was produced using Puzzle, the software for producing a phylogenetic tree, based on these sequences. From the result of alignment and the composite phylogenetic tree, six ancestral mutation, A336F, Y309I, I310L, I321L, A325P and G326S, were predicted using similar procedure as described in Example 1 and 4. The meaning of the notation such as A336F is identical to the meaning described in Example 1 and 4. Among them, since Y309I and I310L, and also A325P and G326S are adjacently located and are located in the same secondary structure, they were considered as a double mutant, respectively. Therefore, Y309/I310L mutation, I312L mutation, A325P/G326S mutation and A336F mutation will be also hereinafter referred to as N1, N2, N3 and N4 mutation, respectively. ^{B15}